DOCUMENT RESUME

ED 306 298 TM 013 176

AUTHOR Beaton, Albert E.
TITLE Sampling Design for the 1990 Trial State Assessment Program.
PUB DATE 3 Apr 89
NOTE 9p.; Paper presented at the Annual Meeting of the American Educational Research Association (San Francisco, CA, March 27-31, 1989).
PUB TYPE Reports - Research/Technical (143) -- Speeches/Conference Papers (150)

EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS Comparative Analysis; Differences; Elementary School Students; Elementary Secondary Education; Mathematics Tests; *National Programs; Pilot Projects; Private Schools; Program Administration; Public Schools; Reading Tests; Sample Size; *Sampling; Science Tests; Secondary School Students; *State Programs; *Testing Programs
IDENTIFIERS *National Assessment of Educational Progress; Stratified Sampling; *Trial State Assessment Program

ABSTRACT

The introduction of the trial state assessment program into the design of the 1990 National Assessment of Educational Progress (NAEP) raises questions about differences across states in sampling and administration practices. In addition, questions about the general approach to comparing state data to national data need discussion. Subject areas covered by the NAEP include reading, mathematics, and science; state programs will only cover mathematics. NAEP samples are selected from both age and grade populations; 9-, 13-, and 17-year-olds and fourth-, eighth-, and twelfth-grade students will be involved. State assessments will only involve eighth-graders. The NAEP will cover private as well as public schools, while the state component will cover only the latter. NAEP will involve a deeply stratified, multi-level sampling plan, with oversampling of minority students and private schools. The NAEP sample will allow regional, but not state-by-state, comparisons. The target for the state-level program includes 2,000 eighth-graders selected from about 100 schools in each state. While special staff are provided for the NAEP, state programs will be conducted by local school staff under rigid guidelines. Special subsampling procedures have been developed for both national and state populations. (TJH)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)
☑ This document has been reproduced as received from the person or organization originating it.
☐ Minor changes have been made to improve reproduction quality.
• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY
Albert E. Beaton
TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

ED306298

Sampling Design for the 1990 Trial State Assessment Program[1]

Albert E. Beaton
Educational Testing Service

April 3, 1989

The introduction of the trial state assessment program into the design of the 1990 National Assessment of Educational Progress (NAEP) has required addressing some rather special technical issues in order not to compromise the validity of either of the assessment programs and to assure that both assessments will be economically feasible. The aim of the trial state assessment program is to compare the students in the many participating states to those in the other participating states and also to the students in the nation as a whole. The technical questions, therefore, have focussed on how the NAEP and the state samples can be assessed and fairly compared while maintaining the integrity of the national assessment. The trial state comparison program cannot have a 'Lake Wobegon' effect in which every state in the Union is reported to be "above average."

Before addressing the comparison of the state and national data, it is important to have an understanding of the differences between the two assessments, and so they will be compared and contrasted in the next sections of this paper. Differences exist not only in the general aims of the two types of assessment but also in the sampling and administrative procedures. After the differences are discussed, the general approach to comparing the state data to the national data will be presented. A more detailed description of the sampling design is available in Beaton (1988).

Assessment Features

The national NAEP has its own important agenda and aims, which are required in the enabling legislation, and cannot be compromised for the trial state comparisons. The national NAEP's most important aim is to report on trends in education in the United States, and thus it must maintain continuity with past assessment practices. To achieve its national goals, the 1990 design of NAEP involves assessing representative samples of students in American schools in the subject areas of reading, mathematics, and science. As in recent years, samples are selected from both age and grade populations, that is, the national sampling design includes

[1]Paper delivered at the annual meeting of the American Educational Research Association, San Francisco, March 29, 1989.

TM013176

BEST COPY AVAILABLE

representative samples of not only 9-, 13-, and 17-year-old students but also of students in the fourth, eighth, and twelfth grades. Both public and private school students are included in the national sample, and in 1990, for the first time, private school students will be over-sampled in order to achieve representative samples of the private school population.

The trial state assessment program has different goals from the national assessment, although the goals overlap. The trial state assessment program in 1990 will assess mathematics only, not three subject areas as in the national. It will assess eighth grade students only, not three grade levels. It will not try to estimate the performance of any age population. It will assess public school students only, not private school students. State participation is voluntary, and the state samples must be large enough for reliable state comparisons.

Assessment Instruments

The instruments of the 1990 national assessment will be, for the most part, BIB spiraled in order to allow each subject area assessment to cover a broad range of topics while keeping the burden on individual students to about one hour. For BIB spiraling, the exercises in each subject area are divided into seven blocks, each of which is expected to be completed by a student in fifteen minutes. The reading, mathematics, and science assessment booklets will each contain

- a five minute block of background and attitude questions that will be asked of all students,
- a five minute block of questions about experiences in the subject area being assessed, and
- three fifteen minute blocks of exercises in a particular subject area.

The seven blocks in each subject area are placed in booklets in a balanced incomplete design as shown in Table 1. Each of the seven blocks is placed in three different booklets in such a way that each pair of blocks occurs in exactly one booklet. To do so, seven booklets are printed for each subject area. Reading, mathematics, and science booklets are then spiraled together in a random sequence and assigned to successive students in an assessment session.

<Insert Table 1 about here>

It should be noted that the national NAEP also has other booklets for other purposes including trend analyses, intercorrelations among subject areas, and measuring mathematical estimation ability using tightly timed stimuli.

The trial state comparison program will use the same BIB-spiraled mathematics booklets that are used for the national NAEP eighth grade

sample. As Mullis (1989) has elaborated, the consensus process by which the 1990 mathematics objectives were determined was conducted under the auspices of the Council of Chief State School Officers with the participation of a very broad spectrum of opinion. The mathematics items have recently been field tested. Many state and local school personnel have had and will continue to have a substantial input into the 1990 mathematics assessment.

Sampling

The sampling for the national assessment will involve a deeply stratified, multilevel sampling plan that first selects primary sampling units (PSUs), then schools within the PSUs, and finally students within the schools. Areas with a high proportion of minority students will be over-sampled in order to assure that a sufficient number are assessed for reliable results. Assessment will be done in two equivalent national samples, one assessed in the winter and the other assessed in the spring. The national NAEP samples have been designed to be large enough for reliable estimates of the performance of various regions of the country (North, Southeast, Central, and West) but not large enough for estimates of student proficiency in individual states. Increasing the national sample sizes to permit state comparisons would be prohibitively expensive, and so another approach to state comparisons had to be developed.

The general NAEP sampling frame would not be particularly efficient for the state assessment sampling since quite good listings of the public schools in each state are available. After verification of the lists, schools will be stratified by one variable such as urbanicity or percent minority and a sample of schools will be selected. Assessment will be done in February of the assessment year. The target for each state sample is 2,000 eighth graders selected from around 100 schools in the state. This sample size will be large enough for reliable estimates of eighth grade proficiency in mathematics for the state as a whole. The sample will also be large enough to report some sub-populations separately (e.g., boys and girls) but not large enough for reliable reporting on other sub-populations (e.g., racial/ethnic groupings and community types) in some states.

Administration

In order to minimize the disruption of the educational programs in the cooperating schools and to attain uniformity of assessment procedures, the national assessment is administered in the schools by a special staff who are supplied by the NAEP contractor or sub-contractor. Hiring, training, and transporting the necessary staff is expensive, but has been an integral part of the NAEP measurement process since its beginning.

Simply extending the national assessment administrative procedures to the trial state comparison samples is simply not economically feasible. Instead, the trial state comparisons will be administered under tightly defined procedures that are implemented by personnel selected by the

participating states. Each participating state will provide a state coordinator and, for each sampled district, a district coordinator. State and district personnel will be trained by the NAEP contractor for field administration (currently Westat, Inc.). The state and district coordinator will train the coordinators in the selected schools.

Comparing State and National Proficiencies

The major differences between the national and state assessments are summarized in Table 2. We first note that the states will all be taking the same assessment booklets and using the same administrative procedures, and thus the results for the different states should be comparable to each other, regardless of the differences between the national sample and the state samples. The differences may, however, affect the comparisons of the states to the nation as a whole. If the differences are not taken into account in the assessment design, there is a possibility that the results will be unacceptable.

<Insert Table 2 about here.>

For example,

- o there is the possibility that only high performing states will participate, in which case all participating states may be above average, an apparent 'Lake Wobegon' effect.

- o the state samples may be more highly motivated than the national. Students in the national assessment are assured that their results will not be reported individually or even at the school or state levels, and they have never been. The state samples are, however, to be reported at the state level for comparison with other states. The newer ambiance around the assessment may make the individual student more highly motivated to perform optimally. Although higher motivation is actually desirable in itself, it may distort valid comparisons with the national data.

- o the uncontrolled supervision of the assessment administration may result in less valid and reliable measurement.

These threats to the validity of the trial state assessment data cannot be safely ignored. For this reason, we have designed special sub-samples for both the national and state designs.

1. The state samples will each be divided into two randomly equivalent halves, one of which will be observed during the assessment and the other not observed. The observer will be employed by the NAEP contractor and instructed to try to correct and report any procedural breaches that he or she may find.

The purpose of randomly dividing the state samples is to estimate the effect, if any, of changing the administrative procedures. Estimating the effect of unobserved administration is essential before introducing the new administrative procedures into future state comparisons. From the reports of the observers, we will be able to judge the adequacy of the training programs and instructions. Although we do not expect any systematic differences between the two samples within a state, we do expect more variance in the unobserved samples if the assessment procedures are less strictly adhered to. The effect of unobserved administration can be estimated by comparing the two state samples.

2. The national sample will be gathered in such a way that a single sub-sample will represent the same population of students as the combined participating state samples. That is, if 40 states participate in the trial state comparisons, then it will be possible to select a sub-sample from the national sample that will represent the aggregate of those 40 states. To do this, the national sampling frame has had to be enlarged.

In principle, the mathematics proficiency of public school eighth graders in the national NAEP sub-sample is the same as that in the combined state samples, except for sampling error which is estimable. However, the responses to individual mathematics items may differ by more than sampling error due to differential motivation or other uncontrolled factors. A substantial difference between the estimated proficiencies of the two samples is clearly unacceptable. However, since the mathematical proficiency of the two samples is in principle identical, the combined state data can be equated to the national data in the construction of the mathematics scales. The steps in equating are discussed by Johnson (1989).

This leaves the concern about a 'Lake Wobegon' effect. There can be no 'Lake Wobegon' effect in national NAEP since it is its own norming sample and only half of the students can be above the median on any scale. Under the plan to equate the combined state data to the national, the same guarantee of only half the students being above the median would hold true if all states were to participate in the trial state comparison. However, all states are not expected to participate and, if only the top performing states were to participate, it is conceivable that all participating states would have average performance scores above the national norm. This situation is unlikely, but not impossible.

However, if only the top performing states do participate, then they <u>should</u> receive higher than average scores, and <u>should not</u> be compared to each other solely in such a way that any high performing state is necessarily listed at the bottom of the participating states. The national sub-sample of participating states makes it possible to show how the group of participating states compares to the nation as a whole and by so doing destroys the logical basis for the "everyone is above average" phenomenon.

References

Beaton, Albert E. (1988) National Assessment of Educational Progress: Design of the 1990 Assessment. Princeton, NJ: Educational Testing Service.

Johnson, Eugene G. (1989) Issues and Procedures in Analyzing the State Assessment Data. Paper given at the annual meeting of the American Educational Research Association, San Francisco, March 29, 1989.

Mullis, Ina V.S. (1989) The 1990 Assessment Instruments Paper given at the annual meeting of the American Educational Research Association, San Francisco, March 29, 1989.

Table 1

1990 NAEP Booklet Design

Booklet Number	General Questionnaire	Specific Questionnaire	Subject Area Blocks First	Second	Third
(Timing)	(5 min.)	(5 min.)	(15 min.)	(15 min.)	(15 min.)
1	yes	yes	Block 1	Block 2	Block 4
2	yes	yes	Block 2	Block 3	Block 5
3	yes	yes	Block 3	Block 4	Block 6
4	yes	yes	Block 4	Block 5	Block 7
5	yes	yes	Block 5	Block 6	Block 1
6	yes	yes	Block 6	Block 7	Block 2
7	yes	yes	Block 7	Block 1	Block 3

The General Questionnaire contains background and attitude questions that are administered to all students.

The Specific Questionnaire contains questions about experiences in the subject area being assessed in the following subject area blocks.

The Subject Area Blocks contain items in a particular subject area (e.g., mathematics). The blocks are assigned to booklets in such a way that each pair of blocks occurs in some booklet.

Table 2

Comparison of National and State Assessments (1990)

	National	State
Subjects assessed	Reading Mathematics Science	Mathematics only
Student populations	Grade 4/age 9 Grade 8/age 13 Grade 12/age 17	Grade 8 only
School populations	Public and Private	Public only
Administration	Contractor staff	State staff
Assessment time	Jan-Mar (1st half) Mar-May (2nd half)	February
BIB spiral booklets	Yes	Yes
Excluded students form	Yes	Yes
Absent student reports	Yes	Yes